Application of modern regression techniques and artificial neural networks on pavement prediction modeling

Y. H. Lee & Y. B. Liu
Department of Civil Engineering, Tamkang University, Taipei, Taiwan

H. W. Ker
Department of International Trade, Chihlee Institute of Technology, Taipei, Taiwan

ABSTRACT: This study strives to illustrate the benefits of incorporating the principles of dimensional analysis, subject-related knowledge, and statistical knowledge into pavement prediction modeling process. Modern regression techniques including local regression and regression splines as well as back propagation neural networks were briefly introduced. Factorial 2-D and 3-D finite element runs and BISAR runs for different pavement systems were conducted to generate the deflection databases for the analysis. The resulting ANN model using all dominating dimensionless parameters was proved to have higher accuracy and require less network training time than the other counterpart using purely input parameters. Increasing the complexity of ANN models does not necessarily improve the modeling statistics. The results also showed that using higher number of neurons and hidden layers sometimes lead to even worse modeling statistics which was an indication of over training and should be avoided. Several local regression models requiring minimal amount of modeling time were also developed using the same databases.

KEY WORDS: Pavement deflection, prediction modeling, dimensional analysis, local regression, artificial neural networks.

1. INTRODUCTION

Predictive models have been widely used in various pavement design procedures, evaluation, rehabilitation, and network management systems. Empirical and mechanistic-empirical approaches using statistical regression techniques have been utilized extensively in predicting extremely complicated pavement responses and performance indicators for more than four decades. Using purely empirical concepts to develop predictive models is not recommended. Lee (1993) proposed a systematic statistical and engineering modeling approach which strongly recommends to incorporate theoretical engineering knowledge, expert experience, heuristics, and statistical data analysis and regression techniques altogether into the framework to develop more mechanistic-based predictive models. In additional to the conventional “parametric” linear and nonlinear regression techniques, several ingenious iterative regression techniques in the area of “robust” and “nonparametric” regressions were also incorporated. The proposed approach has been successfully implemented in the development of many purely empirical predictive models (Lee et al., 1993; Lee & Darter, 1995), purely mechanistic predictive models (Lee & Darter, 1994a; 1994b) as well as the mechanistic-empirical predictive models adopted in the early analyses of LTPP general pavement studies data (Simpson et al., 1993).

Significant progress has been reported in pavement prediction modeling of simulated data using artificial neural networks (ANN). Back propagation networks (BPN) can be taught from one data space to another using representative set of data to be learned. The learning process actually refers to a multi-layered, feed-forward neural network trained by using an error back propagation algorithm or an error minimization technique (Haykin, 1999; Hecht-Nielsen, 1990). Ceylan (2004) conducted a literature search summarizing recent ANN applications in pavement structural evaluation such as backcalculating pavement layer moduli and predicting primary pavement responses (e.g., stress and deflection). As with many ANN applications in the literature (Haussmann et al., 1997; Meier et al., 1997; Ceylan et al., 1998; 1999; 2004; Ceylan & Guclu, 2005), original pertinent input parameters were used to generate the training and testing databases. This approach often requires tremendous amount of time and efforts in network training and testing. To reduce the size of the required factorial databases, researchers sometimes opt to fix certain input parameters to some prescribed values as a special case study, which may result in limiting the inference space of the resulting model.

Nevertheless, some earlier ANN literature has also illustrated that the incorporation of the principles of dimensional analysis lead to significant savings during the training set generation. Ioannides et al. (1996) trained a back propagation neural network (BPN) to determine the in situ load transfer efficiency of rigid pavement joints from Falling Weight Deflectometer (FWD) data. Khazanovich and Roesler (1997) developed an ANN-based backcalculation procedure for composite pavements. The multilayer elastic program DIPLOMAT was used to analyze a three-layer pavement system consisting of an AC surface layer over a PCC slab resting on a Winkler foundation. Ioannides et al. (1999) trained BPN models to predict the critical slab bending stress for loading-only, curling-only, and loading-and-curling cases. BPN predictions were compared against the Westergaard closed-form solutions as well as the statistical regression models developed by Lee and Darter (1994a) using a small set of factorial data with dimensionless mechanistic variables. It was re-emphasized that mature engineering judgment and in-depth understanding of the mechanics of the phenomenon remain the most reliable guides in the formation of the problems to be analyzed.

Attoh-Okine (1994) proposed the use of ANN models in predicting roughness progression of flexible pavements. Although the results were promising, some built-in functions including learning rate and momentum term which form key neural network algorithm were not investigated. Attoh-Okine (1999) used real pavement condition and traffic data and specific architecture to investigate the effect of learning rate and momentum term on BPN models for the prediction of flexible pavement performance. Sorsa et al. (1991) indicated that adding many hidden layers gets the network to learn faster and the mean square error becomes a little smaller, but the generalization ability of the network reduces.

Ripley (1993) discussed many statistical aspects of neural networks and tested it with several benchmark examples against traditional and modern regression techniques, such as generalized discriminant analysis, projection pursuit regression, local regression, tree-based classification, etc. Ripley concluded that in one sense neural networks are little more than non-linear regression and allied optimization methods. "That two-layer networks can approximate arbitrary continuous functions does not change the validity of more direct approximations such as statistical smoothers, which certainly 'learn' very much faster" (Ripley, 1993). Projection pursuit regression highlights the value of differentiated units and other training schemes and offers computation shortcuts through forward and backward selection. Statistical and subject-related knowledge can be used to guide modeling in most real-world problems and so enable much more convincing generalization and explanation, in ways which can never be done by 'black-box' learning systems (Ripley, 1993).

As part of continuous research efforts in pavement design and analysis (Lee et al., 1994a; 1998; 2004), modern regression techniques and artificial neural networks (ANN) are utilized in this study to improve the prediction accuracy of simulated pavement deflections (Wu, 2003; Liu, 2004). Factorial 2-D and 3-D finite element runs and BISAR runs for different pavement systems are conducted to generate the deflection databases for the analysis. This study strives to illustrate the benefits of incorporating the principles of dimensional analysis, subject-related knowledge, and statistical knowledge into prediction modeling process.

2. MODERN REGRESSION TECHNIQUES

2.1 Revised two-step modeling approach using projection pursuit regression

The proper selection of regression techniques is one of the most important factors to the success of prediction modeling. Since most of the regression algorithms currently available do not directly consider interaction effects during the modeling process, the interaction terms must be subjectively determined prior to performing a regression analysis. With the multi-dimensional pavement engineering problems in mind, several unresolved deficiencies are frequently identified in the use of stepwise regression and nonlinear regression. These include problems in the selection of correct functional form, violations of the embedded statistical assumptions, and failure to satisfy some engineering boundary conditions.

The projection pursuit regression (PPR), however, appears to have the most favorable features in handling these problems, which strives to model the response surface (y's) as a sum of nonparametric functions of projections of the predictor variables (x's) through the use of super smoothers. More technical details about the development process, the application, and the demonstration on modeling interactions of the PPR algorithm can be found in the literature (Friedman & Stuetzle, 1981; Friedman, 1984; Mathsoft, Inc. 1997). The S-PLUS statistical package, which has been widely used by statisticians, was selected for the analysis due to the availability of this regression technique.

As a result, a two-step regression analysis procedure was proposed by Lee and Darter (1994b) to better find the correct functional form and to better fit the response surface. With the help of the PPR, a multi-dimensional response surface is broken down into the sum of several smooth projected curves which are graphically representable in two dimensions. Plausible functional forms and applicable boundary conditions may then be easily identified and specified through visual inspection and/or engineering knowledge of physical relationships to model these individual projected curves separately. Traditional parametric regression techniques such as linear, piecewise-linear, and nonlinear regressions are then utilized for these purposes with higher confidence in the parameter estimates.

In this study, regression spline algorithm (Ker, 2002) was adopted in lieu of piecewise-linear regressions at the second step to assure smooth junctions at the change points. A spline function is a piecewise polynomial regression. An *n*-spline function is an *n*-degree polynomial with *n*-1 continuous derivatives at the change points. These change points are called "knots" in spline literature. Spline functions can be viewed as a data-smoothing regression function and/or a way to improve polynomial approximation of regression function. In most cases, a spline can be represented as a linear combination of some basis functions that have polynomial forms. Polynomials can be viewed as a special case of spline with no knots (Smith, 1979). In fitting a spline model, the prediction should be within the data range. Cubic splines with continuous second derivatives at the knots are most commonly used in most applications (Seber & Wild, 1989). Cubic splines are most popular in spline applications because they are of low degree and relatively smooth (assuming continuity restriction up to second derivative only), and possess the power to incorporate several different trends in the range of the data by increasing the number of knots (Smith, 1979).

2.2 Locally-weighted regression (loess) technique

The locally weighted regression (loess) technique is an approach to regression analysis by local fitting developed by Cleveland and Devlin (1988). Cleveland and Grosse (1991) provided computational methods for local regression. A particular data structure called k-d tree is used for partitioning space by recursively cutting cells in half by a hyperplane orthogonal to one of the coordinate axes. The loess approach uses a smoothing technique for fitting a nonlinear curve to the data points locally, so that any point of the curve depends only on the observations at that point and some specified neighboring points. The number of neighbors (k) is specified as the percentage of the total number of points or “span”. Local regression models provide much greater flexibility in fitting a multi-dimensional response surface as a series of many sub-divided regions with single smooth functions of all the predictors. There are no restrictions on the relationships among the predictors.

Figure 1 depicts the concept of loess k-d tree algorithm. This algorithm is available in the S-PLUS statistical package (Mathsoft, Inc., 1997). As currently implemented, locally quadratic models may have at most 4 predictor variables and locally linear models may have at most 15 predictors. The original FORTRAN and C codes for the loess algorithm can also be obtained from the ftp site: “ftp research.att.com.”

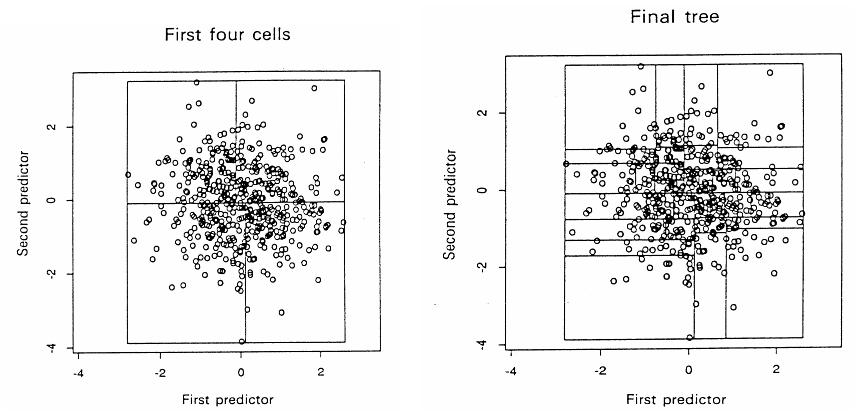

Figure 1. Illustration of loess k-d tree algorithm (Cleveland & Grosse, 1991).

3. ARTIFICIAL NEURAL NETWORKS

Artificial Neural Networks (ANN) provides a flexible way to generalize linear regression functions. They are nonlinear regression models but with so many parameters extremely flexible to approximate any smooth function. The most commonly used rule is the generalized delta rule or back propagation algorithm. Ripley (1993) provided the detail definitions and brief derivation of a back propagation network (BPN). The learning procedure has to select the weights and the biases by presenting the training examples in turn several times, while striving to minimize the total squared error:

$$E = \frac{1}{2}\sum_{p}\left\| y^{p} - c^{p} \right\|^{2} \tag{1}$$

Where y^p is the output for input x^p, and c^p is the target output; the index p runs through the data in the training set. However, the questions of how many layers and how many neurons should be used were treated very lightly in the literature.

A neural network modeling software package called Qnet v2000 for Windows (Vesta Services, Inc. 2000) was adopted for this study. The convergence characteristics of various activation (or transfer) functions including step function, logistic or sigmoid function, hyperbolic tangent function, and radial basis function as shown in Figure 2 will be further investigated (Mehrotra et al., 1997; Smith, 1996).

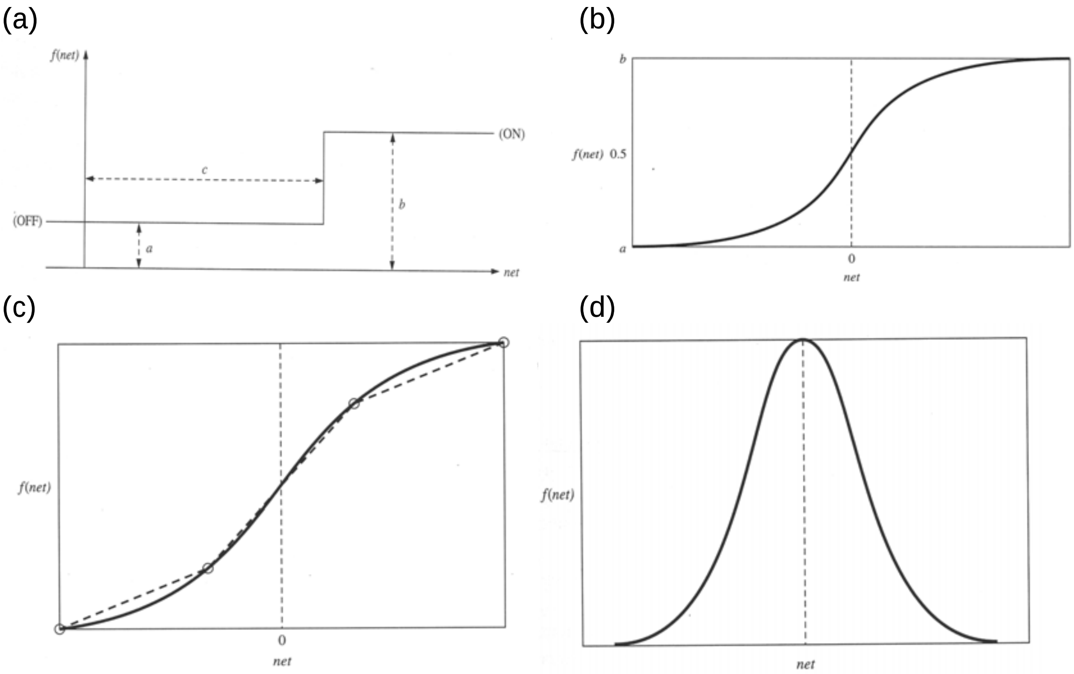

Figure 2. Illustration of various activation (or transfer) functions: (a) step function, (b) logistic or sigmoid function, (c) hyperbolic tangent function, and (d) radial basis function.

4. APPLICATIONS OF ARTIFICIAL NEURAL NETWORKS AND MODERN REGRESSION TECHNIQUES

4.1 Rigid pavement deflection prediction models of infinite slab size

Based on the principles of dimensional analysis, Ioannides et al. (1989) indicated that the structural responses of a rigid pavement such as or the dimensionless deflection parameter ($\delta k\ell^2/P$) are dominated by the following four dimensionless variables: the normalized load radius (a/ℓ), the normalized finite slab length (L/ℓ), the normalized finite slab width (W/ℓ), and the normalized radial distance (r/ℓ) for 2-D FEM analysis. In which δ is the deflection, [L]; k is the modulus of subgrade reaction, [FL^{-3}]; P is the single wheel load, [F]; $\ell=(E*h^3/(12*(1-\mu^2)*k))^{0.25}$ is the radius of relative stiffness of the slab-subgrade system [L]; E is the modulus of the concrete slab, [FL^{-2}]; h is the thickness of the slab, [L]; μ is the Poisson's ratio. Note that primary dimension for force is represented by [F], and length is represented by [L]. To illustrate the benefits of incorporating the principles of dimensional analysis into the modeling process, the following case studies were conducted:

4.1.1 ANN models

For an infinite single slab resting on a Winkler foundation under interior loading condition, factorial ILLI-SLAB runs were conducted based on the following input parameters: single wheel load P=40 kN (9,000 lbs); tire pressure p=0.62 MPa (90 psi); modulus of the concrete slab E= 13.78~48.23 GPa (2~7 Mpsi); modulus of subgrade reaction k=13.5~175.5 MN/m^3 (50~650 pci); and slab thickness h= 15.2~76.2 cm (6~30 in.). These input parameters were such selected to cover wider ranges of practical cases. The dependent variable is the deflection δ and the explanatory variables are *E, k, h,* and *r*. The resulting deflection database consists of 12,329 data points, in which 11,329 observations were randomly selected for actual training and the remaining 1,000 data points was used to monitor the training process. Step activation function was first tried with extreme difficulty in achieving convergence. Subsequently, sigmoid activation function was chosen for the modeling process. The summary statistics of the NET1 model is shown in Table 1. Note that since certain input parameters were fixed to some prescribed values to reduce the size of the required factorial database, the applicability of this special case study is rather limited.

Table 1. Comparison of two different ANN models

ANN Type	NET1	NET2
Outputs		R
Inputs	E, k, h, r	a/ℓ, r/ℓ
Data Points	Training: 11,329 Monitoring: 1,000	Training: 394 Monitoring: 100
Hidden Layer(s)	2	1
Neurons in Each Hidden Layer	12-12	6
Learning Cycle	30,000	10,000
Learning Rate	0.5	0.1
Modeling Time	6 hrs 43 min.	42 min.
RMS	Training: 0.00290 Monitoring: 0.00420	Training: 0.00377 Monitoring: 0.00360
Coefficient of Determination, R^2	0.999	0.9999

Alternatively, the aforementioned factorial ILLI-SLAB runs may be generalized based on the following dimensionless parameters: a/ℓ=0.05~0.4 (step by 0.01) and r/ℓ ranges from 0 to 3.2 determined by automatic mesh generation. To simulate infinite slab size conditions, L/ℓ and W/ℓ were greater than or equal to 8. Thus, a 2-D rigid pavement deflection database with 494 data points was obtained (Liu, 2004). The dependent variable is the deflection ratio (R) defined as the ratio of the deflection at any radial distance to the resulting maximum deflection. In which 394 data points were used for actual ANN training and the remaining 100 observations were used to monitor the training process. The convergence characteristics of various activation functions were investigated. As shown in Figure 3(a), it was noted that sigmoid activation function has better convergence characteristics than hyperbolic tangent function. Using a single hidden layer with only 5 neurons, sigmoid function completed 10,000 training cycles in 35 minutes whereas hyperbolic tangent function needed 60 minutes, although the resulting root mean squared errors (RMS) had no much difference. Radial basis activation function was also tried with extreme difficulty in achieving convergence. In addition, increasing the number of neurons during the network training process does not necessarily improve the modeling statistics. On the contrarily, as shown in Table 2 and Figure 3(b) the resulting RMS and training time increased while increasing the number of neurons in the hidden layer. Since the model with only six neurons had the lowest RMS, it was chosen as the proposed model (NET2) as summarized in Table 1. It was also concluded that with the incorporation of dimensional analysis in the modeling process, the requirements on database generation and network training time could be greatly reduced.

Table 2. Summary statistics of different ANN models

ANN Type	Number of Neurons in the Hidden Layer					
	5	6	7	8	9	10
Training RMS	0.00416	0.00377	0.00524	0.00569	0.00554	0.00520
Monitoring RMS	0.00384	0.00360	0.00492	0.00529	0.00520	0.00490
R-Squared	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
Training Time	35 min.	42 min.	52 min.	60 min.	67 min.	82 min.

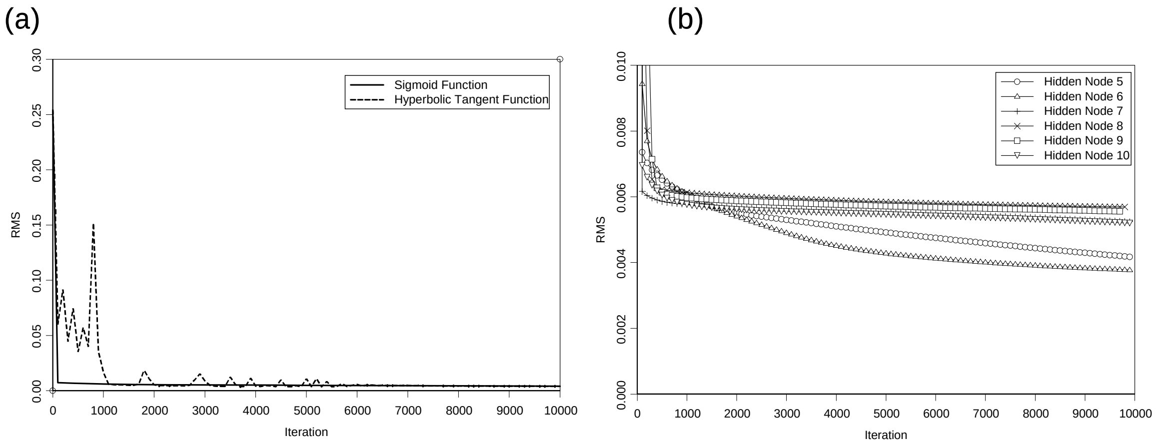

Figure 3. Comparison of convergence characteristics: (a) due to different activation functions; (b) due to different number of neurons in the hidden layer.

4.2 Rigid pavement deflection predictions of finite slab size

To further investigate the convergence characteristics of different ANN models and to illustrate the possibility of over training, the following case studies were conducted.

4.2.1 ANN models

Similarly, for a finite single slab resting on a Winkler foundation under interior loading condition, factorial ILLI-SLAB runs were conducted based on the following input parameters: a/ℓ=0.05~0.4, L/ℓ= 2~7, W/ℓ=2~7, and r/ℓ ranges from 0 to 3.2 determined by automatic mesh generation. A 2-D rigid pavement deflection database with 2,227 data points was obtained (Liu, 2004). The dependent variable is the deflection ratio (R) defined as the ratio of the deflection at any radial distance to the resulting maximum deflection. The explanatory variables are the following dimensionless variables: a/ℓ, L/ℓ, W/ℓ, and r/ℓ.

In which 2,027 data points were randomly selected for actual ANN training and the remaining 200 observations were used to monitor the training process. Similarly, it was noted that sigmoid activation function has better convergence characteristics than hyperbolic tangent function. Using a single hidden layer with only 8 neurons and learning rate = 0.01, sigmoid function completed 30,000 training cycles in 11 minutes whereas hyperbolic tangent function needed 20 minutes, although the resulting root mean squared errors (RMS) had no much difference. Radial basis activation function was also tried, but extreme difficulties were encountered in achieving convergence. By increasing the number the hidden layers from 1 to 2 and the number of neurons from 8 to 13 during the network training process, the resulting RMS and training time are summarized in Table 3. The convergence characteristics of different ANN models with 8 neurons in the first hidden layer were shown in Figure 4. The ANN model with 8 neurons in the first hidden layer and 1 neuron in the second hidden layer was chosen as the proposed model due to its relatively small RMS. The results also showed that more complicated ANN models using higher number of hidden layers and neurons sometimes lead to even worse modeling statistics which was an indication of over training and should be avoided.

Table 3. Summary statistics of ANN models with different number of layers and neurons

ANN Type		Summary Statistics	Number of Neurons in the First Hidden Layer					
			8	9	10	11	12	13
Number of Neurons in the Second Hidden Layer	0	Training RMS	0.01037	0.00965	0.00974	0.00782	0.00887	0.00925
		Monitoring RMS	0.01007	0.00966	0.01046	0.00785	0.00923	0.01011
		R-Squared	0.9988	0.9989	0.9989	0.9993	0.9991	0.9989
		Training Time	11 min.	12 min.	13 min.	15 min.	16 min.	17 min.
	1	Training RMS	0.00552	0.00550	0.00562	0.00553	0.00602	0.00539
		Monitoring RMS	0.00565	0.00594	0.00518	0.00513	0.00562	0.00568
		R-Squared	0.9997	0.9997	0.9996	0.9997	0.9997	0.9997
		Training Time	11 min.	12 min.	13 min.	16 min.	17 min.	21 min.
	2	Training RMS	0.00714	0.00620	0.00563	0.00668	0.01102	0.00589
		Monitoring RMS	0.00711	0.00604	0.00581	0.00613	0.01028	0.00713
		R-Squared	0.9994	0.9996	0.9994	0.9995	0.9988	0.9995
		Training Time	13 min.	13 min.	15 min.	18 min.	19 min.	22 min.
	3	Training RMS	0.00599	0.00751	0.00581	0.00817	0.00981	0.01008
		Monitoring RMS	0.00549	0.00831	0.00588	0.00864	0.00904	0.01061
		R-Squared	0.9997	0.9988	0.9991	0.9988	0.9979	0.9978
		Training Time	15 min.	16 min.	17 min.	19 min.	21 min.	24 min.
	4	Training RMS	0.00570	0.00748	0.00558	0.01005	0.00673	0.00671
		Monitoring RMS	0.00569	0.00803	0.00559	0.01074	0.00656	0.00726
		R-Squared	0.9993	0.9988	0.9994	0.9978	0.9978	0.9991
		Training Time	18 min.	18 min.	19 min.	22 min.	24 min.	25 min.

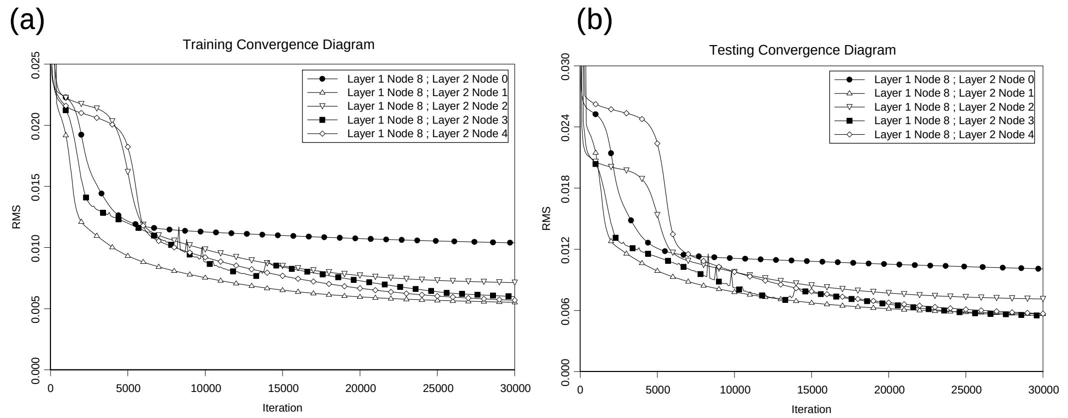

Figure 4. Comparison of convergence characteristics: (a) training data; (b) testing data.

4.2.2 Loess models

Several S-PLUS trials of local regressions were conducted using the same database. The response variable was chosen as the deflection ratio (R) and the explanatory variables were a/ℓ, L/ℓ, W/ℓ, and r/ℓ. The resulting loess model was easily obtained requiring minimal amount of modeling time, in which the smoothing parameter “span” was chosen as 0.1, whereas the “cell” argument was chosen as 0.01. The following regression statistics were obtained: the number of observations = 2,227; equivalent number of parameters = 31.9; residual standard error = 0.006376; and multiple R-squared = 1. The resulting errors were still relatively small even when the proposed loess model was quite simple.

4.3 Three-dimensional rigid pavement deflection predictions

With the introduction of three-dimensional (3-D, ABAQUS) FEM (Hibbitt et al., 2000) and all the promising features reported in the literature, its applications on pavement engineering become inevitable (Wu, 2003). Based on the principles of dimensional analysis, Ioannides and Salsilli-Murua (1989) indicated that the dimensionless deflection parameter ($\delta k\ell^2/P$) is only a function of a/ℓ, L/ℓ, and W/ℓ for 2-D FEM analysis. Extreme difficulties were encountered while using only these three dimensionless variables (a/ℓ, L/ℓ, W/ℓ) to determine $\delta k\ell^2/P$ for 3-D FEM analysis. Subsequently, an additional dominating dimensionless variable (h/a) defined as the ratio of slab thickness (h) and load radius (a) was identified to account for the theoretical differences between 2-D and 3-D FEM analyses (Lee et al., 2004). A series of 3-D FEM factorial runs was conducted for a single squared slab resting on a Winkler foundation under interior loading condition with the following dimensionless parameters: a/ℓ=0.05, 0.1~0.5 (step by 0.1); L/ℓ=2~8 (step by 1); $W/\ell=L/\ell$; and h/a=0.5~6 (step by 0.5). These ranges were carefully selected to cover a very wide range of highway and airfield rigid pavement conditions. An automated analysis program was developed using the Visual Basic software package (Microsoft, 1998) to automatically construct FEM models, generate the input files, conduct the runs, as well as summarize the results to avoid untraced human errors. A 3-D rigid pavement deflection database with 504 data points was obtained (Liu, 2004).

4.3.1 ANN models

In which, 404 observations were randomly chosen for actual training and the remaining 100 data points were used for monitoring the training process. Deflection ratio (R) defined as the ratio of 3-D FEM results to Westergaard solutions was treated as the response variable. Sigmoid activation function was chosen in this case study. The learning rate was set as 0.02 for the cases analyzed. In the first ANN model (NET1), no transformation was made on the response variable. As shown in Table 4 and in Figure 5, the modeling statistics and the convergence characteristics of the NET1 model were satisfactory.

Table 4. Comparison of two different ANN models

ANN Type	NET1	NET2
Outputs	R	1/R
Inputs	a/ℓ, L/ℓ, h/a	a/ℓ, L/ℓ, h/a
Hidden Layer(s)	2	2
Neurons in Each Hidden Layer	10-4	10-4
Learning Cycle	30,000	30,000
RMS	Training: 0.00989 Monitoring: 0.01019	Training: 0.00539 Monitoring: 0.00478
Coefficient of Determination, R^2	0.9988	0.9999

Nevertheless, it is worth mentioning that since Westergaard's closed-from deflection is very small for thicker pavements or larger load sizes (larger h/a and a/ℓ), the resulting 3-D FEM deflections can be several times of the theoretical solutions due to possible compression across the slab thickness. Since the resulting 3-D FEM deflections are always higher than the Westergaard solutions, the reciprocal of the deflection ratio (1/R) always ranges from 0 to 1. Wu (2003) has illustrated that using 1/R as the response variable lead to better physical meanings (or interpretations) of the proposed PPR model. With the incorporation of subject-related knowledge into the modeling process, it was shown that smaller root mean squared errors (RMS) and higher coefficient of determination (R^2) have been achieved in the NET2 model, although the convergence rate was slightly slower.

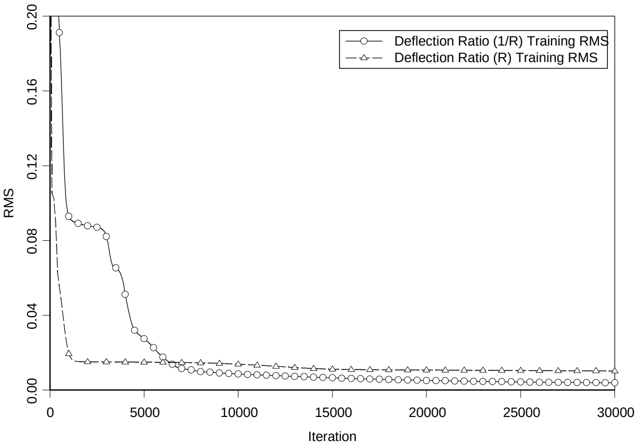

Figure 5. Comparison of the convergence results of two trained ANN modles.

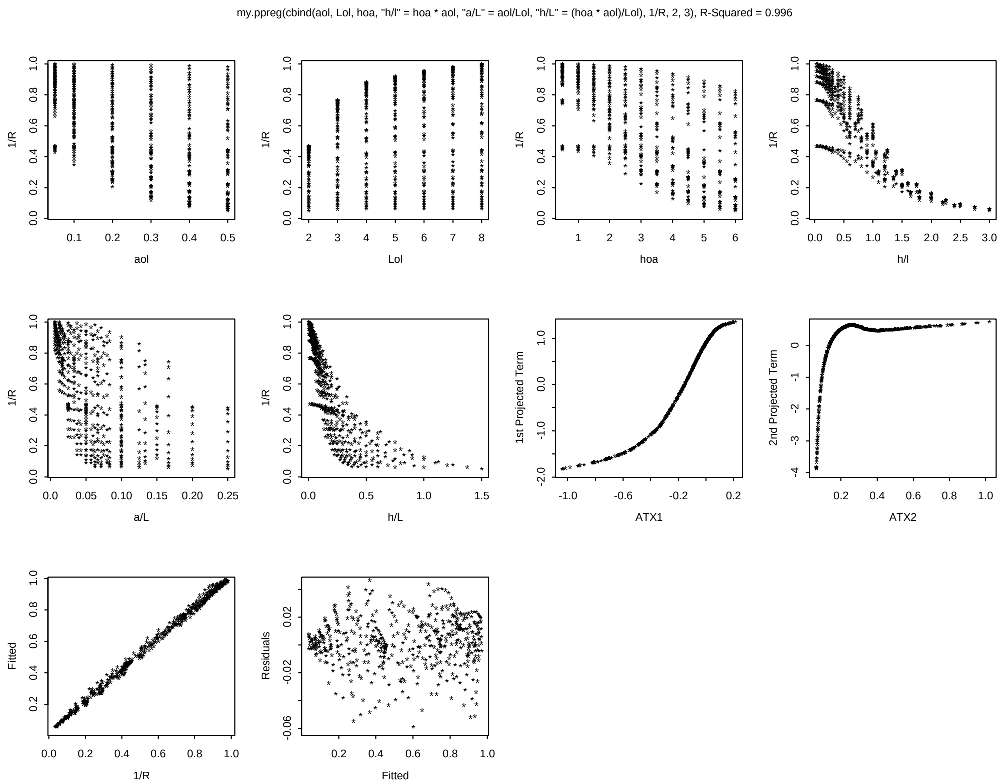

Figure 6. Proposed PPR model for the 3-D deflection database

4.3.2 Revised two-step modeling approach using PPR and regression splines

To facilitate future possible applications of the 3-D rigid pavement deflection database, the following predictive model as shown in Figure 6 was developed using projection pursuit regression technique (Lee & Darter, 1994b; Friedman & Stuetzle, 1981). The response variable was chosen as the reciprocal of the deflection ratio (1/R) and the explanatory variables were *a/ℓ*, *L/ℓ*, *h/a*, and their variations. Regression spline algorithm was adopted in lieu of piecewise-linear regressions at the second step to assure smooth junctions at the change points. Consequently, the coefficient of determination (R^2) was slightly reduced from 0.996 to 0.9942 as the expense of this smoothing. The tentative predictive model and its regression statistics are as follows: (In which, N is the number of observations and SEE is the standard error of the estimation.)

$$1/R = 0.54008 + 0.29653\ \Phi_1 + 0.09667\ \Phi_2$$
$$\Phi_1 = 1.28770 + 10.04098(\mathrm{A1}) + 11.76579(\mathrm{A1})^2 + 4.88399(\mathrm{A1})^3$$
$$-16.23312(A1 > -0.3) * (A1 + 0.3)^3 - 20.95209(A1 > -0.1) * (A1 + 0.1)^3$$
$$\Phi_2 = -11.68986 + 180.53050(\mathrm{A2}) - 895.12897(\mathrm{A2})^2 + 1482.36407(\mathrm{A2})^3$$
$$-1468.55407(A2 > 0.2) * (A2 - 0.2)^3 - 16.40377(A2 > 0.4) * (A2 - 0.4)^3 \qquad (2)$$
$$A1 = \ 0.42473\mathrm{x1} + 0.01922\mathrm{x2} - 0.00925\mathrm{x3} - 0.49378\mathrm{x4} - 0.60805\mathrm{x5} + 0.45343\mathrm{x6}$$
$$\mathrm{A2} = -\ 0.28347\mathrm{x1} + 0.03160\mathrm{x2} + 0.00071\mathrm{x3} + 0.37804\mathrm{x4} + 0.53626\mathrm{x5} - 0.69868\mathrm{x6}$$
$$X = [\mathrm{x1}, \mathrm{x2}, \mathrm{x3}, \mathrm{x4}, \mathrm{x5}, \mathrm{x6}] = \left[\frac{\mathrm{a}}{\ell}\ ,\ \frac{L}{\ell}\ ,\ \frac{h}{a}\ ,\ \frac{h}{a} * \frac{\mathrm{a}}{\ell}\ ,\ \frac{\mathrm{a}}{\ell} / \frac{L}{\ell}\ ,\ \frac{h}{a} * \frac{\mathrm{a}}{\ell} / \frac{L}{\ell}\right]$$
$$\mathrm{Statistics: N = 504, R^2 = 0.9942, SEE = 0.02241}$$

4.3.3 Loess models

Several S-PLUS trials of local regressions were conducted using the same database. Again, the response variable was chosen as the reciprocal of the deflection ratio (1/R) and the explanatory variables were a/ℓ, L/ℓ, and h/a. The resulting loess model was easily obtained at a greatly reduced amount of modeling time, in which the smoothing parameter “span” was chosen as 0.1, whereas the “cell” argument was chosen as 0.1. The following regression statistics were obtained: the number of observations = 504; equivalent number of parameters = 56.6; residual standard error = 0.004784; and multiple R-squared = 1.

4.4 Flexible pavement deflection predictions

Based on the multi-layer elastic theory and the principles of dimensional analysis, the following dominating dimensionless variables were identified for a three-layer pavement system: E_1/E_2, E_2/E_3, h_1/h_2, and a/h_2. In which, a is the radius of the applied load, [L]; h_1 and h_2 are the thickness of the surface and base layers, [L]; E_1, E_2, and E_3 are the Young’s moduli of the surface layer, base layer, and subgrade, respectively, [FL^{-2}]. A series of factorial BISAR runs was conducted with the following ranges to cover most practical pavement data: $0.5 \leqq E_1/E_2 \leqq 170$, $0.5 \leqq E_2/E_3 \leqq 170$, $0.2 \leqq h_1/h_2 \leqq 2.4$, and $0.5 \leqq a/h_2 \leqq 5.0$. A BASIC program written by Dr. Alaeddin Mohseni was used to automatically generate the input files and summarize the results to avoid untraced human errors. A pavement response database including the aforementioned dimensionless variables, deflections at the center of load (D_0), horizontal strain (ε_t) and vertical strain (ε_v) at the bottom of the surface layer was obtained. A training database with 3,600 data points and an independent testing database with 1,728 data points were used in this study (Liu, 2004).

4.4.1 ANN models

The training database was randomly separated into 3,400 data points for actual training and the remaining 200 observations for monitoring the training process. Hyperbolic tangent activation function was chosen in this case study. The learning rate was set as 0.01. At the first trial (NET1) as shown in Table 5, no transformation was made on both explanatory and response variables. Extreme difficulty was encountered in obtaining reasonable convergence.

Based on the basic assumptions of conventional regression techniques that the random errors are mutually uncorrelated and normally distributed with zero mean and constant variance, and additive and independent of the expectation function, it is desirable to check the normality of the response variable. The Box-Cox (1964) transformation procedure was adopted to find the approximate power transformation of the response variable (D_0). As shown in Figure 7(a), the maximum likelihood estimator λ was approximate 0 indicating that a logarithm transformation was appropriate for D_0 (Weisberg, 1985). Figure 7(b) is the normal Q-Q plot which graphically compares the distribution of $\log(D_0)$ to the normal distribution represented by a straight line. This indicates that the logarithm of D_0 is approximate to normally-distributed. In the second trial (NET2), convergence was obtained though the number of learning cycles and modeling time were still very high. The root mean squared (RMS) errors were computed accordingly.

Table 5. Comparison of three different ANN models

ANN Type	NET1	NET2	NET3
Outputs	D_0	$\mathrm{Log}(D_0)$	$\mathrm{Log}(D_0)$
Inputs	E_1/E_2, E_2/E_3, h_1/h_2, a/h_2	E_1/E_2, E_2/E_3, h_1/h_2, a/h_2	$\log(E_1/E_2)$, $\log(E_2/E_3)$, h_1/h_2, a/h_2
Hidden Layer(s)	3	3	2
Neurons in Each Hidden Layer	20-10-5	15-10-5	12-6
Learning Cycle	Cannot converge	200,000	27,000
Modeling Time	> 24 hrs	10 hrs	26 min
RMS	---	Training: 0.0048 Monitoring: 0.0045	Training: 0.0040 Monitoring: 0.0039

(a) (b)

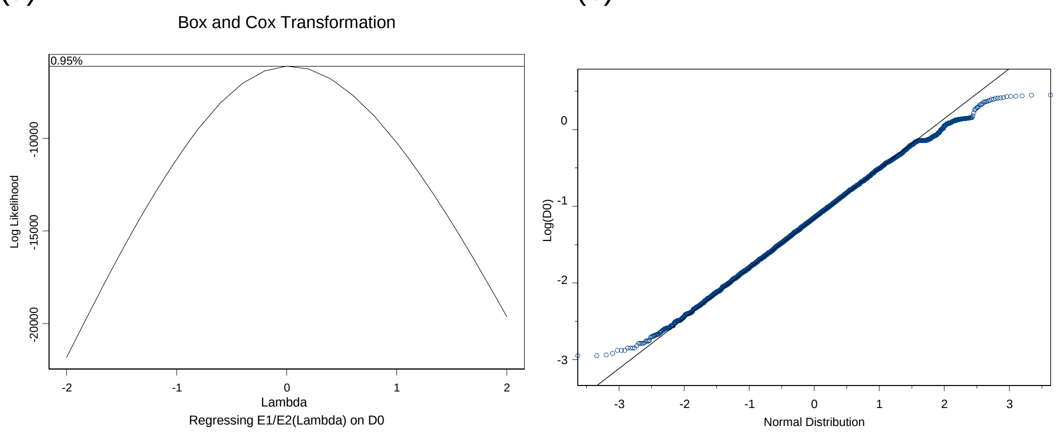

Figure 7. (a) Box-Cox transformation result; and (b) normal Q-Q plot of $\log(D_0)$.

According to general statistical principles or using the alternating conditional expectations (ACE) algorithm (Breiman & Friedman, 1985) together with the Box-Cox power transformation technique proposed by Lee (1993), logarithm transformations of D_0, E_1/E_2, and E_2/E_3 were recommended for NET3 model. As shown in Table 5, with more statistical knowledge incorporated into the ANN modeling process, the resulting ANN model was proved to have higher accuracy and less network training time than the other counterpart using purely input parameters. Figures 8(a) and 8(d) depict the network convergence results for NET2 and NET3 during the training process. The goodness of the prediction of $\log(D_0)$ and the goodness of the prediction of D_0 for NET2 and NET3 were also provided in Figures 8(b)~8(c) and 8(e)~8(f) during the testing phase, respectively. With more statistical knowledge incorporated into the modeling process, the resulting ANN model was proved to have higher accuracy and less network training time than the other counterpart using purely input parameters.

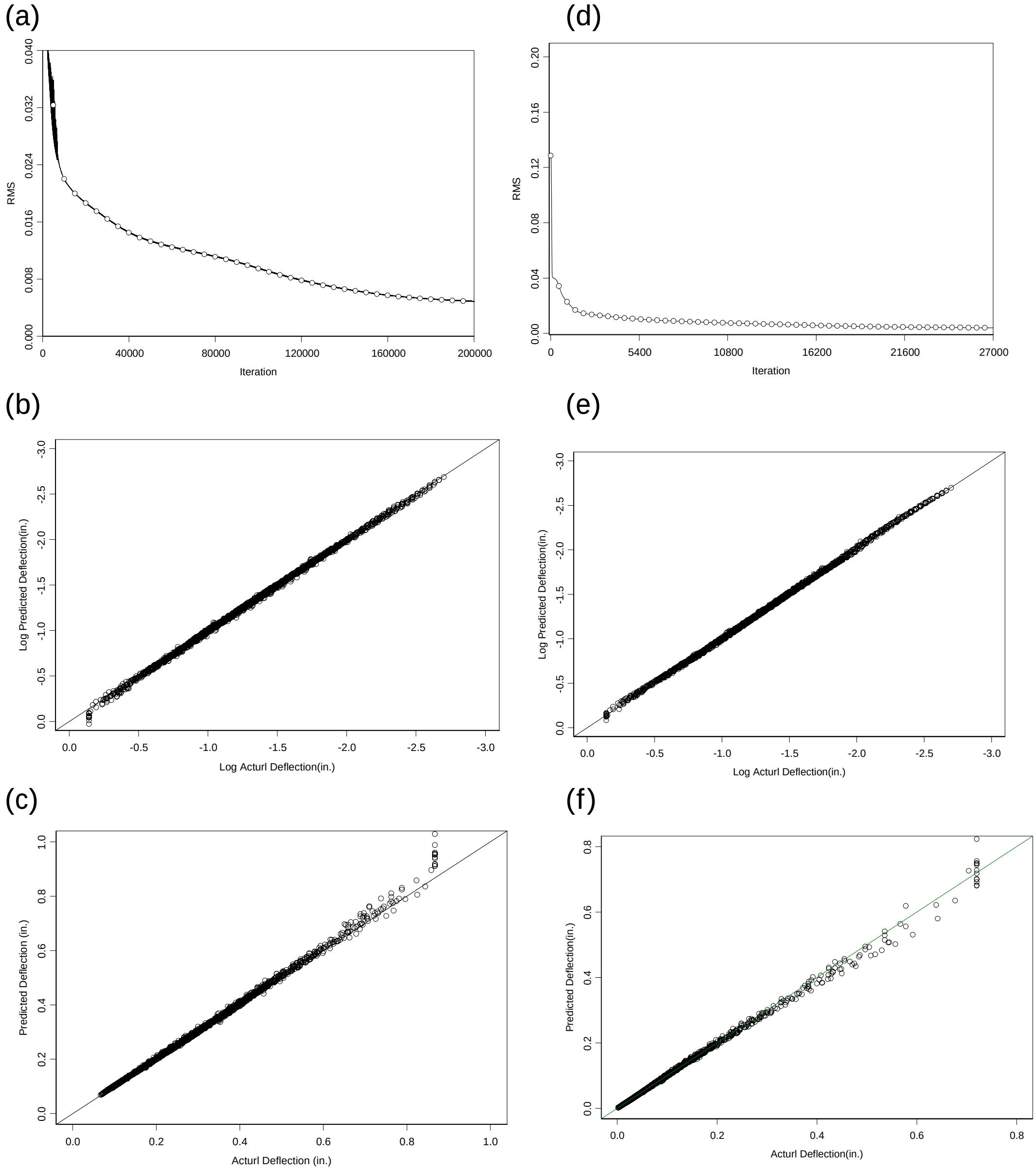

Figure 8. (a) ~ (c) NET2 network convergence results, goodness of the prediction of log(D_0), and prediction of D_0; and (d) ~ (f) for NET3 network, respectively.

4.4.2 Loess models

Several S-PLUS trials of local regressions were conducted using the same training and testing databases. Again, the logarithm transformations of D_0, E_1/E_2, and E_2/E_3 were adopted here. The response variable is log(D_0) and the explanatory variables are log(E_1/E_2), log(E_2/E_3), h_1/h_2, and a/h_2. The resulting loess model was obtained at a greatly reduced amount of modeling time, in which the smoothing parameter "span" was chosen as 0.1, whereas the "cell" argument was chosen as 0.1. The following regression statistics were obtained: number of observations = 3,600; equivalent number of parameters = 31.9; residual standard error = 0.02792; and multiple R-squared = 1. The goodness of the prediction of log (D_0) and D_0 were presented in Figures 9(a) and 9(b), respectively.

The resulting loess model was compared to the aforementioned NET2 and NET3 models for the goodness of D_0 predictions during the testing phase. Reasonable good predictions can be achieved using both ANN and modern regression techniques.

(a) (b)

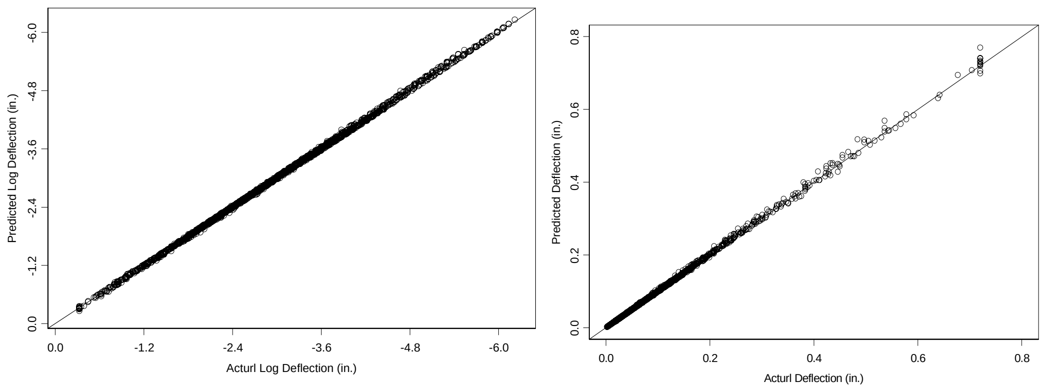

Figure 9. Local regression model: (a) goodness of the prediction of log (D_0) and (b) goodness of the prediction of D_0.

5. CONCLUDING REMARKS

Several case studies were conducted to illustrate the benefits of incorporating the principles of dimensional analysis, subject-related knowledge, and statistical knowledge into pavement prediction modeling process. The resulting ANN model using all dominating dimensionless parameters was proved to have higher accuracy and require less network training time than the other counterpart using purely input parameters. Increasing the complexity of ANN models does not necessarily improve the modeling statistics. The results also showed that using higher number of neurons and hidden layers sometimes lead to even worse modeling statistics which was an indication of over training and should be avoided. Several local regression models requiring minimal amount of modeling time were also developed using the same databases. The resulting loess model was compared to the aforementioned ANN models for the goodness of predictions. Reasonable good predictions can be achieved using both ANN and modern regression techniques. Statistical and subject-related knowledge can be used to guide modeling in most real-world problems and so enable much more convincing generalization and explanation, in ways which can never be done by ‘black-box’ learning systems (Ripley, 1993).

6. REFERENCES

Attoh-Okine, Nii O. (1994). “Predicting Roughness Progression in Flexible Pavements Using Artificial Neural Networks.” *Proceedings*, *Third International Conference on Managing Pavements*, Vol. 1, San Antonio, TX, 55-62.

Attoh-Okine, Nii O. (1999). “Analysis of Learning Rate and Momentum Term in Backpropagation Neural Network Algorithm Trained to Predict Pavement Performance.” *Advances in Engineering Software*, v 30, n 4, 291-302.

Box, G. E. P. and Cox, D. R. (1964). "An Analysis of Transformations (With Discussion)." *Journal of the Royal Statistical Society*, B, Vol. 26, 211-246.

Breiman, L., and Friedman, J. H. (1985). "Estimating Optimal Transformations for Multiple Regression and Correlation (With Discussion)." *Journal of the American Statistical Association*, Vol. 80, 580-619.

Ceylan, H. (2004). “Use of Artificial Neural Networks for the Analysis & Design of Concrete Pavement Systems.” *CD-ROM, the 5th CROW Workshop on Concrete Pavements*, Istanbul, Turkey, March 31-April 3, 2004.

Ceylan, H., and Guclu, A. (2005). “Effects of Combined Temperature and Gear Loading on the Response of Concrete Airfield Pavements Serving the Airbus A380-800 Aircraft.” *CD-ROM,* *the 1st European Airport Pavement Workshop*, Amsterdam, the Netherlands, May 11-12, 2005.

Ceylan, H., Tutumluer, E., and Barenberg, E. J. (1998). "Artificial Neural Networks as Design Tools in Concrete Airfield Pavement Design." *Proceedings, ASCE 25th International Air Transportation Conference*, Austin, Texas, pp. 447-465.

Ceylan, H., Tutumluer, E., and Barenberg, E. J. (1999). "Analysis of Concrete Airfield Slabs Serving the Boeing 777 Aircraft Using Artificial Neural Networks." *Journal of the Transportation Research Board, Transportation Research Record 1684*, 110-117.

Ceylan, H., Tutumluer, E., and Barenberg, E. J. (2004). "Artificial Neural Networks for the Analysis of Slabs Under Simultaneous Aircraft and Temperature Loading." Presented at the 15th ASCE Engineering Mechanics Conference at Columbia University, New York, June 4, 2002. In *Recent Advances in Materials Characterization and Modeling of Pavement Systems*, Edited by Tutumluer, E., Najjar, Y. M., and Masad, E., ASCE, 223-238.

Cleveland, W. S., and Devlin, S. J. (1988). "Locally-Weighted Regression: An Approach to Regression Analysis by Local Fitting." *Journal of the American Statistical Association, 83*, 596-610.

Cleveland, W.S., and Grosse, E. (1991). "Computational Methods for Local Regression." *Statistics and Computing, 1*, 47-62.

Friedman, J. H., and Stuetzle, W. (1981). "Projection Pursuit Regression." *Journal of the American Statistical Association*, 76, 817-823.

Friedman, J. H. (1984). *SMART User's Guide*, Technical Report No. 1, LCS, Department of Statistics, Stanford University, Stanford, CA.

Haussmann, L. D., Tutumluer, E., and Barenberg, E. J. (1997). "Neural Network Algorithms for the Correction of Concerete Slab Stresses from Linear Elastic Layered Porgrams." *Transportaion Research Record 1568*, 44-51

Haykin, S. *Neural Networks: A Comprehensive Foundation*. Prentice-Hall, Inc., New Jersey, 1994.

Hecht-Nielsen, R. *Neurocomputing*. Addison-Wesley, New York, 1990.

Hibbitt, Karlsson, and Sorensen (2000). *ABAQUS/Standard User's Manual*. Vol. I & Ⅱ.

Ioannides, A. M., Alexander, D. R., Hammons, M. I., and Davis, C. M. (1996). "Application of Artificial Neural Networks to Concrete Pavement Joint Evaluation." *Transportation Research Record 1540*, Transportation Research Borad, Washington, D.C., 56-64.

Ioannides, A. M., Barenberg, E. J., and Lary, J. A. (1989). "Interpretation of Falling Weight Deflectometer Results Using Principles of Dimensional Analysis." *Proceedings, Fourth International Conference on Concrete Pavement Design and Rehabilitation*, Purdue University.

Ioannides, A. M., Davis, C. M., and Weber, C. M. (1999). "Westergaard Curling Solution Reconsidered." *Transportation Research Record 1684*, Transportation Research Borad, Washington, D.C., 61-70.

Ioannides, A. M. and Salsilli-Murua, R. A. (1989). "Temperature Curling in Rigid Pavements: An Application of Dimensional Analysis." *Transportation Research Record* 1227, 1-11.

Ker, H. W. (2002). *Application of Regression Spline in Multilevel Longitudinal Modeling*. Ph.D. Dissertation, University of Illinois, Urbana.

Khazanovich, L., and Roesler, J. (1997). "DIPLOBACK: Neural-Network-Based Backcalculation Program for Composite Pavements." *Transportation Research Record 1570*, 1997, 143-150.

Lee, Y. H. (1993). *Development of Pavement Prediction Models*. Ph.D. Dissertation, University of Illinois, Urbana.

Lee,Y. H., and Darter, M. I. (1994a). "Loading and Curling Stress Models for Concrete Pavement Design." *Transportation Research Record 1449*, Transportation Research Board, National Research Council, Washington, D.C., 101-113.

Lee, Y. H., and Darter, M. I. (1994b). "New Predictive Modeling Techniques for Pavements."

Transportation Research Record 1449, Transportation Research Board, National Research Council, Washington, D.C., 234-245.
Lee, Y. H., and Darter, M. I. (1995). “Development of Performance Prediction Models for Illinois Continuously Reinforced Concrete Pavements.” *Transportation Research Record 1505*, Transportation Research Board, Washington, D.C., 1995, 75-84.
Lee, Y. H., Lee, C. T., and Bair, J. H. (1998). “Modified Deflection Ratio Procedures for Backcalculation of Concrete Pavements.” In *Airport facilities: Innovations for the Next Century; Proc., 25th International Air Transportation Conference*, ASCE, M. T. McNerney, eds., Austin, Texas, June 14-17, 480-495.
Lee, Y. H., Mohseni, A. and Darter, M. I. (1993). “Simplified Pavement Performance Models.” *Transportation Research Record 1397*, Transportation Research Board, National Research Council, Washington, D.C., 7-14.
Lee, Y. H., Wu, H. T., and Yen, S. T. (2004). “Parameter Studies on Three-Dimensional Finite Element Analysis of Rigid Pavements.” *CD-ROM, the 5th CROW Workshop on Concrete Pavements*, Istanbul, Turkey, March 31-April 3, 2004.
Liu, Y. B. (2004). *Application of Modern Regression Techniques and Neural Network on Rigid Pavement Backcalculation*. M. S. Thesis, Department of Civil Engineering, Tamkang University, Taipei, Taiwan (In Chinese).
Mathsoft, Inc. (1997). *S-PLUS for Windows (Ver. 4.0) User's Manual, Reference Manual, and Guide to Statistics*.
Mehrotra, K., Mohan, C. K., and Ranka, S. (1997). *Elements of Artificial Neural Networks*. Massachusetts Institute of Technology.
Ripley, B. D. (1993). “Statistical Aspects of Neural Networks.” In *Networks and Chaos - Statistical and Probabilistic Aspects*, Barndorff-Nielsen, O. E., Jensen, J. L., and Kendall, W. S., eds., Chapman & Hall, London, 41-123.
Seber, G. A. F., and Wild, C. J. (1989). *Nonlinear Regression*. New York: Wiley.
Simpson, A. L., Rauhut, J. B., Jordahl, P. R., Owusu-Antwi, E., Darter, M. I., Ahmad, R., Pendleton, O. J., and Lee, Y. H. (1993). *Early Analyses of LTPP General Pavement Studies data*, *Volume 3 - Sensitivity Analyses for Selected Pavement Distresses*. Strategic Highway Research Program, National Research Council, Washington, D.C.
Smith, P. L. (1979). Spline as a Useful and Convenient Statistical Tool. *The American Statistician*, 33, 57-62.
Smith, M. (1996). *Neural Networks for Statistical Modeling*. International Thompson Computer Press.
Sorsa, T., Koivo, H. N., Koivisto, H. (1991). “Neural Nerworks in Process Fault Diagnosis.” *IEEE Transactions on Systems*, *Man and Cybernetics*, 21(4): 815-825.
Vesta Services, Inc. (2000). *Qnet v2000 for Windows*, http://www.qnetv2k.com/.
Weisberg, S. (1985). *Applied Linear Regression*. Second Edition, Weiley Series in Probability and Mathematical Statistics, John Wiley & Sons, Inc.
Wu, H.-T. (2003). *Parameter Analysis and Verification for Jointed Concrete Pavements*. M.S. Thesis, Tamkang University, Tamsui, Taiwan, January 2003. (In Chinese)